

Evaluation of TWAVEZA Activities: A Proposal

31 August 2011

Introduction

In this proposal we describe an approach to evaluation of the activities of TWAVEZA in Tanzania. TWAVEZA is an NGO employing a wide range of activities that should eventually lead to public action aimed at improved public service provision, particularly in the sectors of health, education and water. The type of activities employed by TWAVEZA and the many ways in which these activities could lead to the envisaged outcomes would seem to call for a qualitative, mainly descriptive approach to evaluation: TWAVEZA stresses the importance of mutual dependence of processes leading to “agency” or public action where quantitative research into the workings of these processes would want to disentangle them in order to study them in isolation.

Nevertheless, this proposal has a strong quantitative component while acknowledging TWAVEZA’s multifaceted activities and ambitions. This is achieved by studying society in many places simultaneously. The inevitable differences between events and outcomes from different places will then allow us to identify the quantitative impact of TWAVEZA’s activities.

The evaluation approach outlined here is innovative in the sense that it has not been tried out before although it builds on earlier work by the principal investigators.¹ This means that some of the evaluation activities will take some time to start up and be consolidated into formal research protocols. The evaluation approach is also risky, requiring more than usual monitoring and frequent consultation between TWAVEZA and the evaluators.

Evaluation Approach

There is a widespread interest in rigorous evaluation of policy interventions in developing countries. Typically such interventions take a form, which makes an experimental setup infeasible or undesirable: a policy may for example be targeted at the poor. The implication is that treatment and control groups differ not just in treatment status, but also in the extent of their poverty and possibly in many other ways. To take this into account requires regression techniques, using data from household surveys.

¹ See Elbers, Chris, Jan Willem Gunning and Kobus de Hoop (2009), ‘Assessing Sector-Wide Programs with Statistical Impact Evaluation: a Methodological Proposal’, *World Development*, vol. 37, 2009, pp. 513-520, and Elbers, Chris and Jan Willem Gunning (2010), *Evaluation of Development Policy: Treatment versus Program Effects*, mimeo.

This approach needs to be modified in the TWaweZA case. TWaweZA is unusual in that it does not directly aim at a goal like better education or health services but tries to achieve such outcomes *indirectly*. Many of its activities are aimed at creating awareness of a particular problem at the national level. TWaweZA may draw attention to a problem (such as the late or non arrival of capitation grants) by organizing a workshop, getting newspaper coverage of the issue, or talking to political leaders. These activities may result in information on the topic being discussed in villages. TWaweZA expects that this in turn will prompt “agency” (or “public action”) and eventually improved outcomes. The proposed evaluation will investigate the various steps in this theory of change. This differs radically from the usual evaluation design, which would focus on the last step, from action (“intervention”) to impact.

The proposed evaluation has five components: (i) a sample survey of households at the beginning of the evaluation, to be repeated in 2013; (ii) high-frequency (‘real time’) monitoring of information and agency at village level; (iii) informal interviews in villages (individual or collective); (iv) econometric analysis of the effects of (local) agency on village level outcomes; (v) analyzing the ‘information echo’ of TWaweZA initiatives at national and village levels.

Before describing these components in detail it is good to take a general look at the evaluation approach. Figure 1 summarizes in a simplified way in which TWaweZA initiatives aimed at national organization and institutions eventually affect the local level.

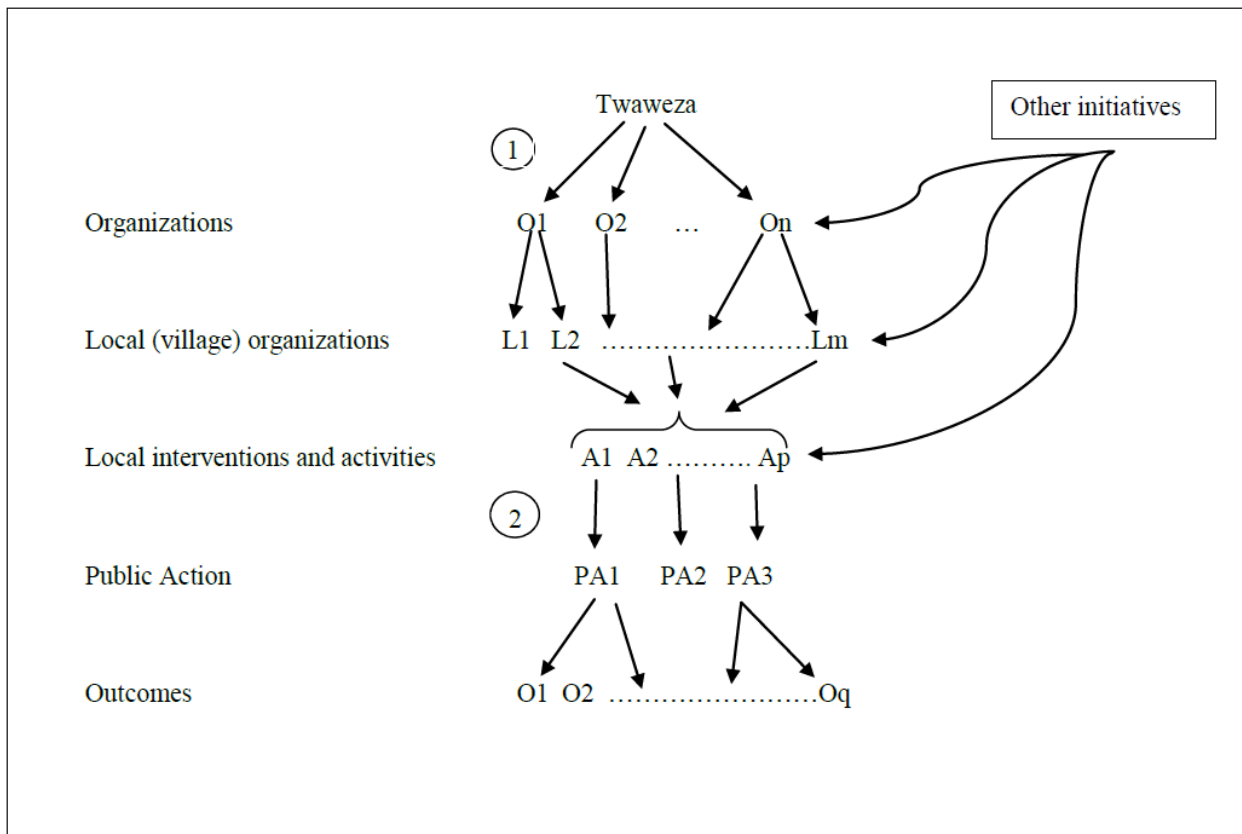


Figure 1: Simplified intervention logic

Twaweza activities are pictured around the circled (1) in the graph. They are aimed at various organizations and institutions. In turn these organizations call on local organizations at the village level, leading to local interventions and activities. Obviously, organizations both national and local, can also be influenced by initiatives *not* originating from TWaweZA. Local interventions (pictured around the circled 2) can lead to public action at the local level and subsequently to outcomes in terms of service delivery or household indicators on health, education and access to safe water facilities. The evaluation design's quantitative part focuses on the elements around circled (2): how local initiatives trigger public action and how public action affects outcomes. However, local initiatives need not have been prompted by TWaweZA activities: they could have many origins. Establishing the relationships around (2) would however provide evidence for a central piece of TWaweZA's theory of change: that local activities and agency can lead to improved public service delivery.

Assuming that it can be shown that relationship (2) works as theorized it remains to establish that TWaweZA, although it may not be responsible for *all* local initiatives, can and does trigger at least some of them. This is the part pictured around circled (1). In the evaluation we will follow the links under (1) qualitatively by a number of case studies and quantitatively by investigating whether there exists an 'information echo' of TWaweZA activities in national and local media. This requires qualitative research similar to expenditure tracking: the researcher attempts to trace the effect of the Twaweza intervention through the organizations Twaweza has talked to, down to the local level. This can obviously not be done exhaustively or very accurately, but it is important to establish plausibility of a channel from the intervention to the local event.

Thus, TWaweZA's theory of change would be 'proved' by positive findings both under (1) and under (2).

We now describe the different components in the evaluation approach in more detail.

(i) Sample survey

The household survey comprises 2500 households from 250 locations (villages, *mtaas*), 10 households per location. The survey is aimed to be representative at various grouping levels of the population. The design of the survey is described in more detail in a separate section below. The survey collects information on outcomes and conditions among households, both at the start of the evaluation (baseline) and three years later. The changes in outcomes in the different locations will be linked statistically to events (or a formal classification of the 'event history') in those locations. The survey is therefore the ultimate yardstick for the impact assessment. The first round of the survey took place early 2011.

(ii) High-frequency monitoring at village level

An innovative element in the evaluation approach is to set up a 'listening device' to record what goes on in the villages during the evaluation period. This will provide necessary information for the event history mentioned above but will also produce valuable information in itself on developments at the grass roots level. For this listening device we will set up a network of local informants who are contacted regularly for a telephone interview about village level events and public action in the location. The challenge is to develop

a semi-structured protocol for the telephone interviews and distill the qualitative information from the interviews into a classification of event histories suitable for econometric analysis.² Developing the listening device and a suitable protocol is expected to take 3-4 months.

We propose to establish an informant in all the 250 villages where TWAVEZA has conducted a baseline survey in 2011. The informant should live in the village and be in a position to monitor what goes on there. He or she would be equipped with a mobile phone and contacted every two to four weeks for a telephone interview. Importantly, the interview should only be structured lightly and the set of topics should not be set in concrete. Rather, it should be adapted to current developments and local interests. The informants should report what people talk about and what actions people engage in. The interviews should lead to a record of events that could lead to the kind of agency aimed at by Twaweza. For example, a local religious leader may give a speech on the poor state of the health clinic. Such events are highly heterogeneous, may occur at unexpected times and are sometimes repeated.

Since TWAVEZA operates mainly nationally it is important for the evaluation that TWAVEZA should not engage in targeting activities specifically at the 250 villages in the sample. More generally, TWAVEZA should act on information from the listening device only after agreement with the researchers in order not to compromise the evaluation.

To keep the 'listening device' outlined above in operation requires highly motivated 'runners' of the system. Moreover, they must be Tanzanian nationals because of language and cultural constraints. We propose to let the system be run by a postdoctoral researcher and four PhD students. The innovative nature of the evaluation approach offers great opportunities for academic achievements: journal publications and material for PhD research. This, as well as sufficient financial compensation, should provide incentives for maintaining the system of informants.

The 250 local informants are supervised by four PhD students in Dodoma. The PhD students conduct the telephone interviews and may occasionally visit the villages and organize training sessions for their informants. During the first two years of their PhD work the interviews and the subsequent analysis will probably take at most half their time. They can use the data (together with the principal investigators) for their thesis. This is expected to provide them with a strong intrinsic motivation to make a success of the data collection through the local informants. In addition the collaboration arrangement between the universities in Dodoma and Amsterdam will enable them to spend time in Amsterdam to follow PhD courses and work on their research papers in a sandwich formula construction. An essential requirement is that the PhD students should be full time available for this research project and not engage in consultancies on other projects. To make this feasible a generous grant will be made available to the students. We consider two economics students and two students from sociology or anthropology for this project.

The PhD students report to a postdoc, also based in Dodoma, who will be working on the project full time for a period of around three and a half years and will be funded accordingly under the project. The postdoc and the PhD students are all envisaged to be Tanzanians. Again, the postdoc is expected to be intrinsically

² For the analysis of qualitative interview data we intend to use a system such as Atlas.ti. See e.g. R.B. Lewis. 1998. *Cultural Anthropology Methods* 10(3): 41-47, and <http://www.atlasti.com/>.

motivated as a principal investigator on the project. The other principal investigators will be Jan Willem Gunning and Chris Elbers and possibly other researchers from Tanzania.

(iii) Informal interviews at local level

The high-frequency village data are to be supplemented by lower frequency information from informally conducted interviews with individuals and groups from the villages. We avoid the term focus group discussion here to stress the informality: the organization of formal group sessions must be avoided. These informal interviews should serve to validate the high-frequency data and check on the quality of the information supplied by the informants. The PhD students are responsible for organizing the interviews, attending them themselves.

(iv) Econometric analysis of outcomes

Information on outcome variables in the main TWAVEZA sectors (health, education and water), on public action-related variables, and on other pre-intervention variables have been collected in the baseline survey and will be collected in a subsequent follow-up survey as mentioned above. TWAVEZA initiatives and agency-related activities take place between the baseline and the later survey. This differs markedly from a traditional treatment/control group design, since TWAVEZA activities are public and in some sense are shared by everyone while it is expected that all locations experience some form of public action.³

The impact of (local) agency can be identified by linking differences in outcomes to differences in agency related activities between locations, accumulated over the evaluation period, 2011-2013. In practice we propose to aggregate agency related activities into a number of classes characterized by likely channels of influence and frequencies of occurrence. Changes in local outcomes can then be related to such location-specific differences in intervention classes. Classes must be chosen so that the effects of interaction between intervention types and frequencies can be identified. A simple example can illustrate this. Consider two interventions, applied at three different intensity levels. This creates a total of 9 different combinations (which obviously would have to be aggregated into larger classes). Regressing outcome differences on class indicators (in the form of 'dummy variables') allows assessment of the differences in effectiveness of combinations of interventions. This strategy can work only if great care is taken to collect intervention histories at the local level.

(v) Information echo of TWAVEZA activities

The final component of our evaluation design investigates how effective TWAVEZA is in raising public awareness at national and local levels. For example, Twaweza introduces a campaign at the national level, e.g. the capitation grant. This should be detectable in national media, in the form of an increase in the number of words and phrases related to the capitation grant. We intend to study the existence of such an 'information echo' of TWAVEZA activities quantitatively, using electronic newspaper scans and data mining of web-sites. Likewise, at the local level, the informants can provide evidence of the degree to which topics stirred up by TWAVEZA reach the village level.

³ A similar evaluation context is studied in Elbers and Gunning (2010).

Neighboring locations

Outcomes and subjective assessment of outcomes may also depend on developments in neighboring locations. This can be tested by including measures of information and activities in neighboring locations among the possible determinants of those outcomes. The sample design discussed below makes this possible. This is a powerful way of studying the importance of spreading information between villages.

Research questions

Summarizing, there are three groups of research questions, concerning, first, whether Twaweza initiatives succeed in triggering local “events”, secondly whether “events” induce public action and, finally, whether public action is effective in terms of outcomes. As indicated above the first group will be analyzed mostly in a qualitative fashion, the other two econometrically. In the analysis of the effects of events and public action (i.e. the second and third groups of questions) we will specifically allow for spillovers between neighboring villages.

A major issue in the analysis concerns “interaction effects”: are particular types of interventions or public action more effective in combination with others? The Twaweza philosophy assumes such nonlinearities and the evaluation allows us to test this by studying the impact of combinations of activities.

Sample Design

The sample consists of 2500 households in 250 villages, 10 households per village. In order to be able to pick up information spillovers between villages we want to sample some of them in “pairs” (2 adjacent villages). The sample consists of 24 of such pairs (48 villages) and 202 unpaired locations. There are three further considerations that have guided the sample design.

First, the sample should be representative at the zonal level. There are seven zones⁴ in Tanzania. Since we have no prior information on inter-zonal differences in the variance of outcomes we propose to sample an equal number of villages from each zone. This amounts to about 36 villages per zone.

Secondly, in all zones some villages have already experienced, prior to the baseline, one Twaweza intervention, namely the learning assessment organized by Uwezo.⁵ Since the Uwezo villages were selected randomly the effect of the intervention can be inferred through comparison with a control group of non-test villages. (Another evaluation team will do this.) For other interventions the fact that the Uwezo test preceded the baseline survey is a serious disadvantage. This was an argument for limiting the number of Uwezo villages in the sample. These two considerations were reconciled by oversampling Uwezo villages but only to a limited extent. A third of the villages were sampled from the Uwezo test villages. By sampling

⁴ These are Central, Eastern, Lake, Northern, Southern Highlands, Southern and Western zones.

⁵ See <http://www.uwezo.net>. The Uwezo villages were selected through stratified sampling: 38 districts were selected randomly from the 133 districts in the nation; within each district 30 urban and rural villages were randomly selected, the share of urban locations being equal to that in the (district) population. From each village the largest government school was picked. Learning tests were administered to the children from 20 randomly households in the village.

Uwezo villages only from the five zones with the largest population (Central, Eastern, Lake, Northern and Southern Highlands zones) the share of Uwezo villages in those zones is 50% which maximizes the statistical power of comparison tests between Uwezo and non-Uwezo villages.

Thirdly, it was considered desirable to have an equal number of rural and urban locations in the sample.

The sample composition by zone, Uwezo status, and rural/urban status is shown in Table 1 below. This composition implies a share of 49% of urban locations (*mtaa*) and 36% percent of Uwezo villages, close to the objectives of 50% and 33% respectively. In the original sample of Uwezo villages the share of urban locations in each district is approximately equal to the urban population share, resulting in an overall 26% share of urban locations. Inflating the 26% to the target of 50% would however have led to excessively high urban shares in some zones (up to 82%). We therefore reduced the number of urban locations in those zones (Eastern and Southern Highlands zones) and increased the number of urban locations in the less urbanized zones (Southern and Western zones) as compensation.

Table 1: Sample Composition					
Zone	Uwezo village	Uwezo locations (mtaa)	Other village	Other urban locations (mtaa)	Total
	Rural	Urban	Rural	Urban	
Central	9	9	9	9	36
Eastern	7	11	7	11	36
Lake	8	10	8	10	36
Northern	9	9	9	9	36
SHL	7	11	7	11	36
Southern			24	11	35
Western			24	11	35
Total	40	50	88	72	250

The pairing operation is limited to the Southern and Western zones where the Uwezo complication does not apply. In each of these zones 12 pairs (24 rural villages) will be sampled. Since the concept of an adjacent location for information transfer makes little sense in an urban context, pairs are not included for the urban locations in the sample.

In each location 10 households are sampled, in two clusters of 5 households. In an urban *mtaa* the clusters are sampled from different sides of the street, in a rural village from two distinct areas.

Deliverables

Regular updates on information from 'listening device'.

The information from the 'listening device' is expected to include information that is highly relevant for current TWaweza activities as well as other government and non-government agencies. Quarterly reports will be written summarizing the main findings. These will be non-technical reports focusing on perceptions of public service delivery and local initiatives and public action.

The researchers, principal investigators, as well as PhD students will make themselves available to TWaweza if participation in media events is desirable and to make contributions to TWaweza's web-site.

Policy briefs

It is expected that the research activities provide useful input for TWaweza's newsletters and policy briefs. Since the independent character of the evaluation must be safeguarded, policy briefs based on the findings should be the responsibility of TWaweza rather than AIID. However, whenever TWaweza considers this desirable the researchers will make relevant material available in a form, which can be readily used by TWaweza staff.

Academic dissemination

The researchers will present the evaluation approach and eventually the results in academic workshops and conferences. This will increase TWaweza's public exposure, but more importantly ensure the credibility of the evaluation approach.

In addition, the researchers intend to publish a number of academic papers related to the evaluation, including PhD theses written by the PhD students. While these papers and PhD theses are not formally deliverables under this proposal, the researchers consider this an important part of their responsibility, if only because the innovative nature of the evaluation needs rigorous peer reviewing. This will require use of data collected under the project. TWaweza will own these data, but will not restrict their use for these purposes in any way.

Evaluation report

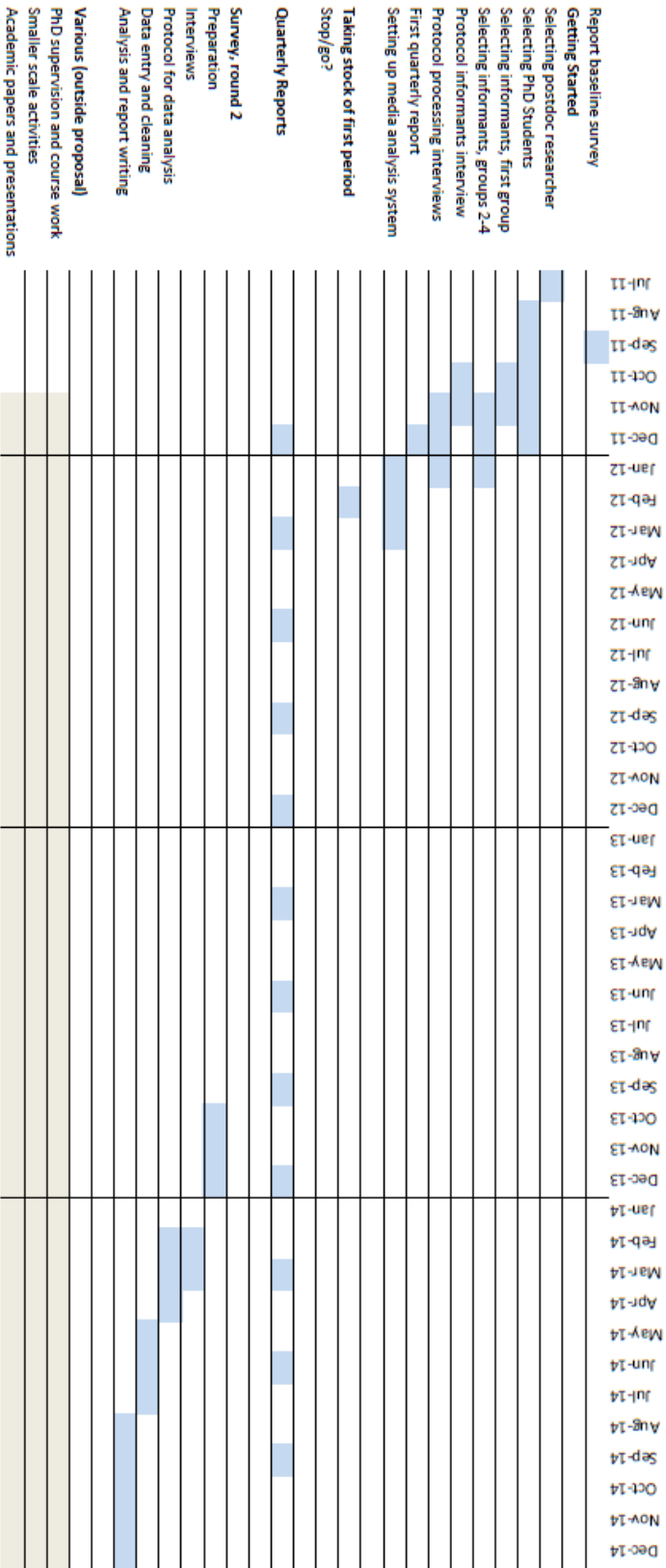
After the second survey round, envisaged for 2013, an extensive report of the evaluation as outlined above will be written by the PIs and PhD students. A draft report will be ready within 3 months after the cleaned survey data will have become available.

- AIID will produce a descriptive report on the baseline survey. This will include tables and graphs to describe the means and distributions of all the key variables. The report will highlight differences between zones, between rural and urban locations etc.
- AIID will be responsible for the second round of the survey. The timing of this data collection is to be agreed with Twaweza.
- AIID will perform the panel data econometric analysis of the various Twaweza activities, using the baseline, the data collected by the informants and the second survey round data. (In accordance with the division of labor agreed with LPT, Uwezo pre-baseline activities will not be analyzed, but information on these interventions will be used as controls in the regressions.)
- all of these activities will be organized in close consultation with Twaweza.

Timing

A schedule for the timing of the various activities is indicated in the Table below. The biggest challenge for the evaluation is setting up the listening device, which is expected to take several months. Early 2012 it should be clear if the evaluation is feasible as planned and this would be a good moment to take stock. Given the novelty of the approach and the many steps involved it would be wise to decide on continuation of the project around February 2012.

With the listening device and the panel of informants in place it is possible to define smaller scale projects or surveys based on the selected locations and the informants. Such projects could be part of the students' PhD research or they could be proposed for other reasons. In order not to interfere with the main evaluation such projects can only be undertaken after consultation and agreement with the principal investigators and TWaweza.



AIID Expertise

AIID's main focus is on the evaluation of programs in the health, education and water/sanitation sectors.⁶ The principal investigators (Chris Elbers and Jan Willem Gunning) have also contributed to methodological innovations in this area. They have a large network of contacts and collaborative relationships in Africa as well as extensive experience with research in the continent. AIID is a multi-disciplinary research network of the two Amsterdam universities involving economists, public health specialists, geographers and political scientists.

The Institute is able to conduct the entire TWAVEZA evaluation in collaboration with researchers in East Africa. The size of the AIID network enables great flexibility in recruitment, including the possible use of PhD students. While the Institute would like to be involved in all aspects of the evaluation proposed here, the principal investigators are particularly interested in the investigation of spillover effects and high-frequency data on interventions.

⁶ See also AIID's website, www.aiid.org.

Budget

	US\$ 000s	
4 PhD students	288	\$ 24,000 each per year; 3 years (starting September 2011)
1 postdoc supervisor	144	full time for 41 months, starting August 2011; in addition to supervision: analysis, academic papers, policy briefs
project admin at IRDP	20	\$ 500 per month, for 40 months
equipment		
phones for enumerators	15	\$ 50 per phone, 300 phones.
equipment IRDP	5	computers, printer/scanner/copier
travel IRDP	47	\$ 15,600 per year, 3 years
stationary IRDP	11	\$ 3,600 per year, 3 year
enumerators		\$ 5 per month; 3 years; 250 enumerators; 30.000 for travel, workshops etc.
	75	
Synnovate round 2	250	
analysis of media impact	50	very rough estimate
travel Amsterdam-Dar	27	16 trips a € 1.200
per diem in Dar	14	\$ 170/day
travel within Tanzania	3	40 days, \$ 85/day
research time of PIs	147	35 days per PI per year a € 500
research assistants and junior researchers		
Amsterdam	60	50 days/year a € 250
baseline questionnaire development		
	9	10.5 days seniors a € 500; 7 days juniors a € 250
AIID overhead	32	5%
contingencies	115	10%
total	1,311	

The budget is based on extensive discussions during a preparatory visit to Dar es Salaam, early 2011. The principal investigators are included at a rate of Euro 500 per day. This is half of their usual rate. This can be justified because they expect to be able to use the results of the analysis in academic work.

Conclusion

The evaluation approach outlined above is ambitious, innovative, and risky. It may therefore be wise to set a particular date for deciding whether the various activities work as planned and if not, whether a major change of plan or even discontinuation of the evaluation should be considered.

Jan Willem Gunning

Chris Elbers

AIID and VU University Amsterdam